



Moving Picture, Audio and Data Coding  
by Artificial Intelligence  
[www.mpai.community](http://www.mpai.community)

## **MPAI Conformance Testing**

### **Multimodal Conversation MPAI-MMC**

**Version 2.1**

#### **WARNING**

Use of the technologies described in this Technical Specification may infringe patents, copyrights or intellectual property rights of MPAI Members or non-members.

MPAI or any of its Members accept no responsibility whatsoever for damages or liability, direct or consequential, which may result from use of this Technical Specification.

Readers are invited to review Annex 2 - Notices and Disclaimers.

# Conformance Testing: Multimodal Conversation (MPAI-MMC) V2.1

|       |   |    |
|-------|---|----|
| 1.    | Introduction .....  | 3  |
| 2.    | Scope of standard .....                                       | 4  |
| 3.    | Terms and definitions .....                                   | 5  |
| 4.    | References .....  | 7  |
| 4.1   | Normative References .....                                    | 7  |
| 4.2   | Informative References .....                                  | 8  |
| 5.    | Conversation with Emotion (CWE) .....                         | 8  |
| 5.1   | Reference Model .....   | 8  |
| 5.2   | AIW Conformance Testing .....                                 | 9  |
| 5.3   | AIM Conformance Testing .....                                 | 9  |
| 5.3.1 | Automatic Speech Recognition .....                            | 9  |
| 5.3.2 | Input Speech Description + PS-Speech Interpretation .....     | 10 |
| 5.3.3 | Input Face Description + PS-Face Interpretation .....         | 10 |
| 5.3.4 | Natural Language Understanding .....                          | 10 |
| 5.3.5 | PS-Text Interpretation .....                                  | 10 |
| 5.3.6 | Multimodal Emotion Fusion .....                               | 10 |
| 5.3.7 | Entity Dialogue Processing .....                              | 11 |
| 5.3.8 | Text-To-Speech (Emotion) .....                                | 11 |
| 5.3.9 | Video Lip Animation .....                                     | 11 |
| 6.    | Multimodal Question Answering (MQA) .....                     | 11 |
| 6.1   | Reference Model .....   | 11 |
| 6.2   | AIW Conformance Testing .....                                 | 12 |
| 6.3   | AIM Conformance Testing .....                                 | 12 |
| 6.3.1 | Visual Scene Description + Visual Object Identification ..... | 13 |
| 6.3.2 | Automatic Speech Recognition .....                            | 13 |
| 6.3.3 | Natural Language Understanding .....                          | 13 |
| 6.3.4 | Question Analysis Module .....                                | 13 |
| 6.3.5 | Answer to Question Module .....                               | 13 |
| 6.3.6 | Text-To-Speech .....  | 14 |
| 7.    | Unidirectional Speech Translation (UST) .....                 | 14 |
| 7.1   | Reference Model .....   | 14 |
| 7.2   | AIW Conformance Testing .....                                 | 14 |
| 7.3   | AIM Conformance Testing .....                                 | 15 |
| 7.3.1 | Speech Recognition .....                                      | 15 |
| 7.3.2 | Text-to-Text Translation .....                                | 15 |
| 7.3.3 | Speech Feature Extraction and Text-To-Speech .....            | 15 |
| 7.3.4 | Text-To-Speech .....  | 15 |
| 8.    | Data Sets .....   | 15 |
| 8.1   | Introduction .....  | 15 |
| 8.2   | Conversation with Emotion .....                               | 16 |
| 8.2.1 | Text .....  | 16 |
| 8.2.2 | Audio and Video Files .....                                   | 17 |
| 8.2.3 | Emotion JSON Files .....                                      | 19 |
| 8.2.4 | Meaning JSON Files .....                                      | 19 |
| 8.3   | Multimodal Question Answering .....                           | 21 |
| 8.3.1 | Text Files .....  | 21 |

|  |   |    |
|--|---|----|
| 8.3.2  | Audio Files .....                       | 22 |
| 8.3.3  | Images .....                            | 22 |
| 8.3.4  | Meaning JSON Files .....                | 22 |
| 8.3.5  | Intention JSON Files .....              | 24 |
| 8.4  | Unidirectional Speech Translation ..... | 25 |
| 8.4.1  | Text Files .....                        | 25 |
| 8.4.2  | Audio Files .....                       | 25 |
| Annex 1 - MPAI-wide terms and definitions .....                                |   | 26 |
| Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative)..... |   | 29 |
| Annex 3 - The Governance of the MPAI Ecosystem (Informative) .....             |   | 31 |

## 1. Introduction

*Technical Specification: Multimodal Conversation (MPAI-MMC)* [2] – in the following also called MPAI-MMC – has been developed MPAI – Moving Picture, Audio, and Data Coding by Artificial Intelligence, the international, unaffiliated, non-profit organisation developing standards for Artificial Intelligence (AI)-based data coding with clear Intellectual Property Rights licensing frameworks [9] in compliance with the rigorous MPAI standards developing process [10] while pursuing the following policies:

1. Be friendly to the AI context but, to the extent possible, agnostic to the technology – AI or Data Processing – used in an implementation.
2. Be attractive to different industries, end users, and regulators.
3. Address three levels of standardisation: data types, components (called AI Modules), configurations of components (called AI Workflows) all exposing standard interfaces with an aggregation level decided by the implementer.
4. Specify the data exchanged by components with a clear semantic to the extent possible.

As manager of the MPAI Ecosystem specified by Governance of MPAI Ecosystem (MPAI-GME) [11], MPAI also ensures that a user of MPAI Technical Specifications can:

1. Operate a Reference Implementation of the Technical Specification, by providing a Reference Software Specification with annexed Reference Software Implementation.
2. Test the Conformance of an implementation of the Technical Specification, by providing the Conformance Testing Specification.
3. Assess the Performance of an implementation of a Technical Specification, by providing the Performance Assessment Specification.
4. Get Conforming Implementations possibly with a Performance Assessment report issued by a trusted source through the MPAI Store.

*Technical Specification: AI Framework (MPAI-AIF)* [1] enables an effective implementation of some of the policies outlined above thanks to its AI Framework (AIF) environment enabling the secure execution of AI Workflows (AIW) constituted by components called AI Modules (AIM).

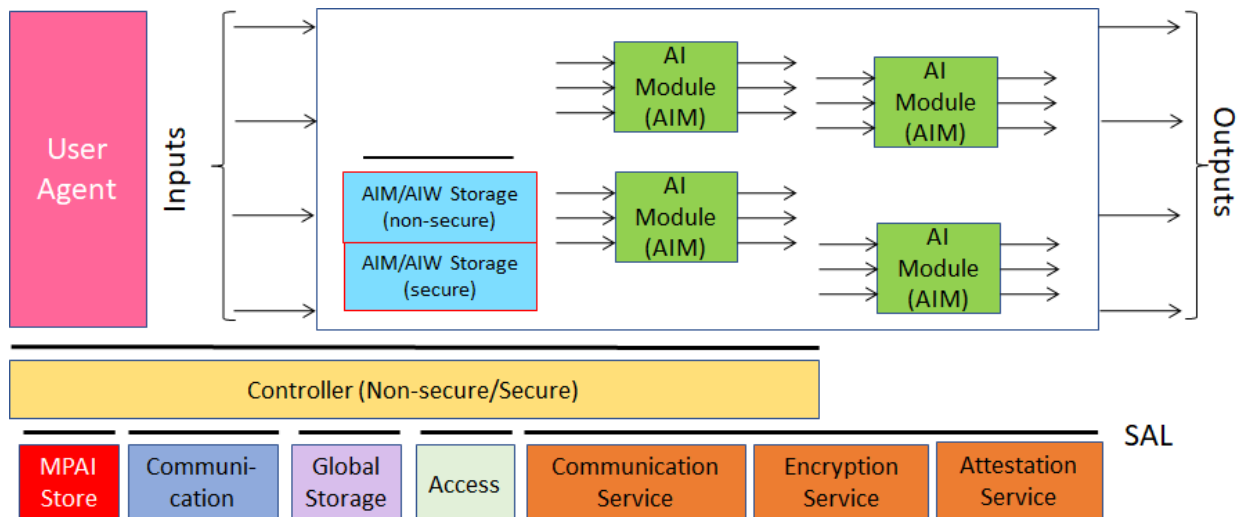


Figure 1 - The AI Framework (MPAI-AIF) V2 Reference Model

With MPAI-AIF, users can execute AI applications having an explicit computing workflow. Component developers can provide components with standard interfaces that have improved performance compared to other implementations.

An AIW and its AIMs may have 3 interoperability levels:

*Level 1* – Implementer-specific and satisfying the MPAI-AIF Standard.

*Level 2* – Specified by an MPAI Application Standard.

*Level 3* – Specified by an MPAI Application Standard and certified by a Performance Assessor.

AI Modules can execute data processing or Artificial Intelligence algorithms and can be implemented in hardware, software, or hybrid hardware/software.

The MPAI-MMC Technical Specification can be implemented in different interoperability types:

1. As a specific data type, as specified in this document.
2. As a specific AIM, as specified in this document.
3. As a specific AIW, as specified in this document.

However, MPAI does not specify the choice of interoperability type, which remains the sole decision of the implementer.

The chapters and the annexes of this *Conformance Testing Specification: Multimodal Conversation (MPAI-MMC)* are Normative unless they are labelled as Informative. Terms beginning with a capital letter are defined in *Table 1* if specific of this MPAI-AIF Conformance Testing Specification, or in *Table 6* if used across MPAI Standards.

## 2. Scope of standard

*Conformance Testing Specification: Multimodal Conversation (MPAI-MMC) V2.1* provides the Procedures, the Conformance Testing Datasets and/or the methods to generate the Conformance Testing Datasets, and the Tools to ascertain whether the Syntax and Semantics of Output Data produced by the specified subsets of AIMs or AIWs conform with *Technical Specification: Multimodal Conversation (MPAI-MMC) V2.1* [2].

This document does not provide tools to Assess the Performance of an Implementation as this function is to be covered by *Performance Assessment Specification: Multimodal Conversation (MPAI-MMC) V2.1* or subsequent versions.

The Version of this MPAI-MMC Conformance Testing covers the following Use Cases of Version 2.1:

1. “Conversation with Emotion” (MMC-CWE), supporting audio-visual conversation with a machine making itself perceptible by a synthetic voice and an animated face.
2. “Multimodal Question Answering” (MMC-MQA), supporting request for information about a displayed object.
3. “Unidirectional Speech Translation” (MMC-UST) supporting conversational translation applications.

The Tools provided for MMC-UST can also be used to Test the Conformance of:

- a. “Bidirectional Speech Translation” (BST).
- b. “One-to-Many Speech Translation” (MST).

This version of *Conformance Testing Specification: MPAI Multimodal Conversation (MPAI-MMC) V2.1* has been developed by the *MPAI Multimodal Conversation Development Committee (MMC-DC)*. MPAI may decide to produce new Versions of this Conformance Testing Specification.

### 3. Terms and definitions

Terms beginning with a capital letter have the meaning defined in *Table 1*. Terms beginning with a small letter have the meaning commonly defined for the context in which they are used. For instance, *Table 1* defines *Object* and *Scene* but does not define *object* and *scene*.

A dash “-” preceding a Term in *Table 1* indicates the following readings according to the font:

1. Normal font: the Term in the table without a dash and preceding the one with a dash should be read before that Term. For example, “Avatar” and “- Model” will yield "Avatar Model.”
2. *Italic* font: the Term in *Table 1* without a dash and preceding the one with a dash should be read after that Term. For example, “Avatar” and “- Portable” will yield "Portable Avatar.”

*Table 1 – Table of terms and definitions*

| <b>Term</b>      | <b>Definition</b>  |
|------------------|--|
| Attitude         |  |
| - <i>Social</i>  | The coded representation of the internal state related to the way a human or avatar intends to position vis-à-vis the Environment or subsets of it, e.g., “Respectful”, “Confrontational”, “Soothing”. |
| - <i>Spatial</i> | Position and Orientation and their velocities and accelerations of an Audio and Visual Object in a Virtual Environment.  |
| Audio            | Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 32, and non-linear and linear quantisation.                     |
| - Object         | Coded representation of Audio information with its metadata. An Audio Object can be a combination of Audio Objects.  |
| - Scene          | The Audio Objects of an Environment with Object location metadata.   |

|                              |  |
|------------------------------|--|
| Audio-Visual Object          | Coded representation of Audio-Visual information with its metadata. An Audio-Visual Object can be a combination of Audio-Visual Objects.   |
| Audio-Visual Scene           | (AV Scene) The Audio-Visual Objects of an Environment with Object location metadata.   |
| Avatar                       | An animated 3D object representing a real or fictitious person in a Virtual Space.   |
| - Model                      | An inanimate avatar exposing interfaces enabling animation.  |
| Cognitive State              | The coded representation of the internal state reflecting the way a human or avatar understands the Environment, such as “Confused”, “Dubious”, “Convinced”.   |
| Colour (of speech)           | The timber of an identifiable voice independent of a current Personal Status and language.   |
| Connected Autonomous Vehicle | A vehicle able to autonomously reach an assigned geographical position by: <ol style="list-style-type: none"> <li>1. Understanding human utterances.</li> <li>2. Planning a route.</li> <li>3. Sensing and interpreting the Environment.</li> <li>4. Exchanging information with other CAV.</li> <li>5. Acting on the CAV’s motion actuation subsystem.</li> </ol> |
| Context                      | Additional information about a communication emitted by an Entity, such as language, culture etc..   |
| Data                         | Information in digital form.   |
| - Format                     | The standard digital representation of Data.   |
| - Type                       | An instance of Data with a specific Data Format.   |
| Descriptor                   | Coded representation of text, audio, speech, or visual feature.  |
| Digital Representation       | Data corresponding to and representing a real entity.  |
| Emotion                      | The coded representation of the internal state resulting from the interaction of a human or avatar with the Environment or subsets of it, such as “Angry”, “Sad”, “Determined”.  |
| Entity                       | A real or Digital Human  |
| Environment                  | A Virtual Space containing a Scene.  |
| Face                         | The portion of a 2D or 3D digital representation corresponding to the face of a human.   |
| Factor                       | One of Emotion, Cognitive State and Attitude.  |
| Gesture                      | A movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance.  |
| Grade                        | The intensity of a Factor.   |
| Human                        | A human being in a real space.   |
| - <i>Digital</i>             | A Digitised or a Virtual Human in a Virtual Space.   |
| - <i>Digitised</i>           | An Object in a Virtual Space that has the appearance of a specific human when rendered.  |
| - <i>Virtual</i>             | An Object in a Virtual Space created by a computer that has a human appearance when rendered but is not a Digitised Human.   |
| Identifier                   | The label uniquely associated with a human or an avatar or an object.  |
| Instance                     | An element of a set of entities – Objects, users etc. – belonging to some levels in a hierarchical classification (taxonomy).  |
| Intention                    | The result of analysis of the goal of an input question.   |
| Manifestation                | The manner of showing the Personal Status, or a subset of it, in any one of Speech, Face, and Gesture.   |

|                    |   |
|--------------------|---|
| Meaning            | Information extracted from Text such as syntactic and semantic information, Personal Status, and other information, such as an Object Identifier.   |
| Modality           | One of Text, Speech, Face, or Gesture.  |
| Object Descriptors | Attribute of the coded representation of an object in a Scene, including its Spatial Attitude.  |
| Orientation        | The set of the 3 roll, pitch, yaw angles indicating the rotation around the principal axis (x) of an Object, its y axis having an angle of 90° counterclockwise (right-to-left) with the x axis and its z axis pointing up toward the viewer. |
| Personal Status    | The ensemble of information internal to a person, including Emotion, Cognitive State, and Attitude.   |
| Portable Avatar    | A Data Type representing an Avatar and its Context.   |
| Pitch              | The fundamental frequency of Speech. Pitch is the attribute that makes it possible to judge sounds as "higher" and "lower."   |
| Point of View      | The Spatial Attitude of a human or avatar looking at an Environment.  |
| Position           | The 3 coordinates (x,y,z) of a representative point of an object in the Real and Virtual Space.   |
| Refined Text       | The Text resulting from the analysis of the Text produced by Automatic Speech Recognition made by Natural Language Understanding.   |
| Scene              | A structured composition of Objects.  |
| Speech             | Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with a number of bits/sample of 8, 16 and 24, and non-linear and linear quantisation.   |
| - Features         | Aspects of a speech segment that enable its description and reproduction, e.g., degree of vocal tension, Pitch, etc., and that can be automatically recognised and extracted for speech synthesis or other related purposes.                  |
| - Rate             | The number of Speech Units per second.  |
| - Unit             | Phoneme, syllable, or word as a segment of Speech.  |
| Summary            | An abridged outline of the content of the utterance(s) of one or more Users possibly including their Personal Statuses.   |
| Text               | A sequence of characters drawn from a finite alphabet.  |
| Visual Object      | Coded representation of Visual information with its metadata. A Video Object can be a combination of Video Objects.   |
| Vocal Gesture      | Utterance, such as cough, laugh, hesitation, etc. Lexical elements are excluded.  |

## 4. References

### 4.1 Normative References

MPAI-MMC normatively references the following documents:

1. MPAI; Technical Specification: [Artificial Intelligence Framework \(MPAI-AIF\) V2.0](#).
2. MPAI; Technical Specification: [Multimodal Conversation \(MPAI-MMC\) V2.1](#).
3. MPAI; Technical Specification: [Object and Scene Description \(MPAI-OSD\) V1.1](#).
4. ISO 639; Codes for the Representation of Names of Languages – Part 1: Alpha-2 Code.
5. ISO/IEC 10646; Information technology – Universal Coded Character Set.
6. ISO/IEC 14496-12:2015; Information technology – Coding of audio-visual objects – Part 12: ISO base media file format

7. ISO/IEC 14496-10:2022; Information technology – Coding of audio-visual objects – Part 10: Advanced video coding.
8. ITU-R; Long-form file format for the international exchange of audio programme materials with metadata; BS.2088-1 (10/2019)  
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000001.shtml>.

## 4.2 Informative References

The following informative references may be useful to clarify aspects of this Technical Specification

9. MPAI; [The MPAI Statutes](#).
10. MPAI; [The MPAI Patent Policy](#).
11. MPAI; [Technical Specification: Governance of the MPAI Ecosystem V1.1](#), 2021.
12. Framework Licence: [Multimodal Conversation \(MPAI-MMC\)](#).
13. MPAI; [Introduction to MPAI-MMC V1](#).

## 5. Conversation with Emotion (CWE)

### 5.1 Reference Model

Figure 2 reproduced the Conversation with Emotion Reference Model from [2].

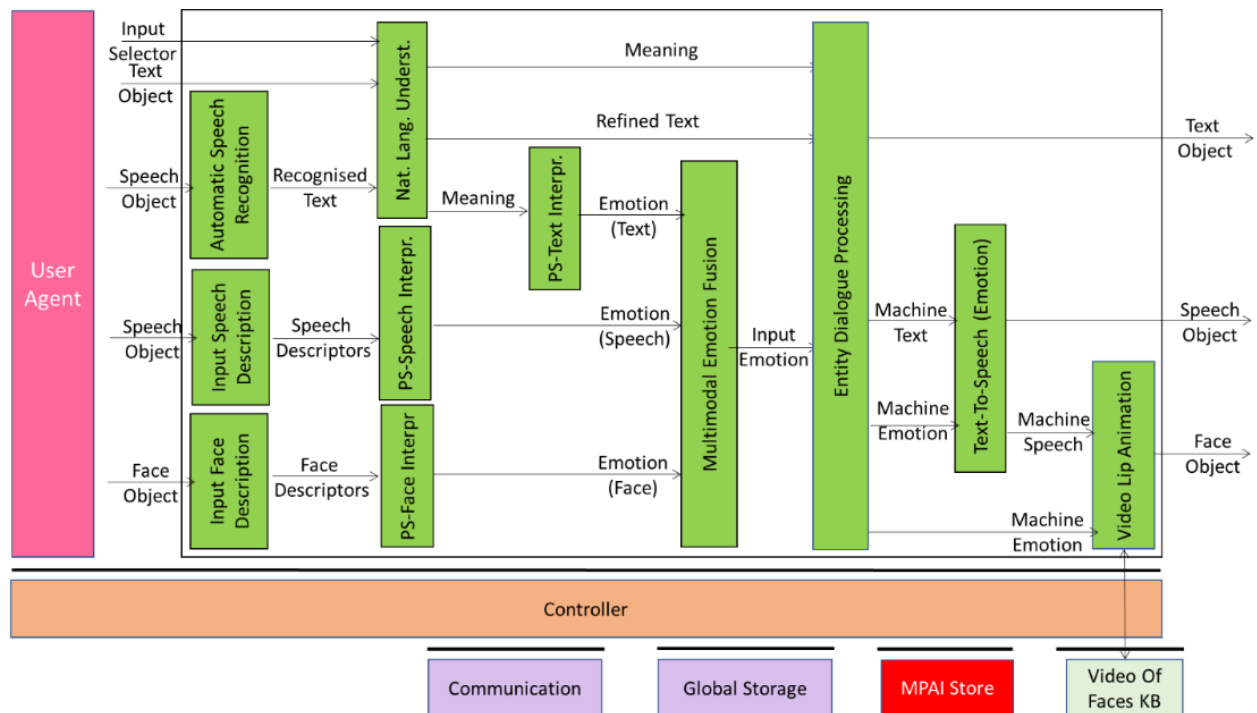


Figure 2 - Conversation with Emotion Reference Model

**Important note.** This Conformance Testing Specification does not provide methods and datasets to Test the Conformance of:

1. Input Speech Description and PS-Speech Interpretation, but only the Composite AIM combining the two.
2. Input Face Description and PS-Face Interpretation, but only their Composite AIM combining the two.



## 5.2 AIW Conformance Testing

Table 2 gives the AIMs and the input/output data of the entire MMC-CWE AI Workflow.

*Table 2 - MMC-CWE AIMs and their I/O data*

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                               |
|--------------------|------------------|---|
| Input Selector     | Binary data      | For Input Selector=0 and Input Selector=1                           |
| Text Object        | Unicode          | All Text files of 8.2.1   |
| Speech Object      | .wav             | All Speech files of 8.2.2   |
| Face Object        | AVC              | All Video files of 8.2.2  |
| <b>Output Data</b> | <b>Data Type</b> | <b>Conformance Test</b>   |
| Machine Text       | Unicode          | All Text files shall conform with [5]                               |
| Machine Speech     | .wav             | All Speech files shall conform with the file format [8]             |
| Machine Video      | AVC              | All Video files shall conform with the file format specified in [7] |

## 5.3 AIM Conformance Testing

Table 3 gives the AIMs and the input/output data of MMC-CWE.

*Table 3 - MMC- CWE AIMs and their I/O data*

| <b>AIM</b>   | <b>Input Data</b>                                     | <b>Output Data</b>              |
|--|---|---------------------------------|
| <b>Automatic Speech Recognition</b>                        | Input Speech  | Recognised Text                 |
| <b>Input Speech Description + PS-Speech Interpretation</b> | Input Speech  | Emotion (Speech)                |
| <b>Input Face Description + PS-Face Interpretation</b>     | Input Video   | Emotion (Face)                  |
| <b>Natural Language Understanding</b>                      | Input Selection<br>Text Object<br>Recognised Text     | Meaning<br>Refined Text         |
| <b>PS-Text Interpretation</b>                              | Meaning   | Emotion (Text)                  |
| <b>Multimodal Emotion Fusion</b>                           | Emotion (Text)<br>Emotion (Speech)<br>Emotion (Video) | Final Emotion                   |
| <b>Entity Dialogue Processing</b>                          | Meaning<br>Recognised Text<br>Input Emotion           | Machine Text<br>Machine Emotion |
| <b>Text-To-Speech (Emotion)</b>                            | Machine Text<br>Machine Emotion                       | Speech Object                   |
| <b>Video Lip Animation</b>                                 | Machine Speech<br>Machine Emotion<br>Video of Face    | Face Object                     |

The Datasets of input and output data of the AIMs are specified in the following paragraphs.

### 5.3.1 Automatic Speech Recognition

| <b>Input Data</b> | <b>Data Format</b> | <b>Input Conformance Testing Data</b> |
|-------------------|--------------------|---------------------------------------|
| Speech Object     | .wav               | All Speech files of 8.2.2             |

| <b>Output Data</b> | <b>Data Format</b> | <b>Conformance Testing criteria</b>                      |
|--------------------|--------------------|--|
| Recognised Text    | Unicode            | Output Text files shall include Text conforming with [5] |

### 5.3.2 Input Speech Description + PS-Speech Interpretation

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                                |
|--------------------|------------------|--|
| Speech Object      | .wav             | All Speech files of 8.2.2  |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>   |
| Emotion (Speech)   | JSON             | All JSON files shall validate against <a href="#">Emotion</a> Schema |

emotion\_Name and emotion\_SetName must be present in the output JSON file of Emotion. The value of either of the two may be null.

### 5.3.3 Input Face Description + PS-Face Interpretation

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                               |
|--------------------|------------------|---|
| Face Object        | AVC              | All Video files of 8.2.2  |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>  |
| Emotion (Face)     | JSON             | All JSON File shall validate against <a href="#">Emotion</a> Schema |

emotion\_Name and emotion\_SetName must be present in the output JSON file of Emotion. The value of either of the two may be null.

### 5.3.4 Natural Language Understanding

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                                |
|--------------------|------------------|--|
| Input Selector     | Binary data      | For Input Selector=0 and Input Selector=1                            |
| Text Object        | Unicode          | All Text files of 8.2.1  |
| Recognised Text    | Unicode          | All Text files of 8.2.1  |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>   |
| Meaning            | JSON             | All JSON files shall validate against <a href="#">Meaning</a> Schema |
| Refined Text       | Unicode          | All Text files shall conform with [5]                                |

The four taggings: POS\_tagging, NE\_tagging, dependency\_tagging, and SRL\_tagging must be present in the output JSON file of Meaning. Any of the four tagging values may be null.

### 5.3.5 PS-Text Interpretation

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                               |
|--------------------|------------------|---|
| Meaning            | JSON             | All JSON Meaning files of 8.2.4                                     |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>  |
| Emotion (Text)     | JSON             | All JSON File shall validate against <a href="#">Emotion</a> Schema |

The two attributes emotion\_Name and emotion\_SetName must be present in the output JSON file of Emotion. The value of either of the two attributes may be null.

### 5.3.6 Multimodal Emotion Fusion

| <b>Input Data</b> | <b>Data Type</b> | <b>Input Conformance Testing Data</b>       |
|-------------------|------------------|---|
| Emotion (Text)    | JSON             | All JSON Emotion (Text) JSON files of 8.2.3 |
| Emotion (Speech)  | JSON             | All Emotion (Speech) JSON files of 8.2.3    |
| Emotion (Video)   | JSON             | All Emotion (Face) JSON files of 8.2.3      |

| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>  |
|--------------------|------------------|---|
| Input Emotion      | JSON             | JSON File shall validate against <a href="#">Emotion</a> Schema |

The two attributes emotion\_Name and emotion\_SetName must be present in the output JSON file of Emotion. The value of either of the two attributes may be null.

### 5.3.7 Entity Dialogue Processing

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                           |
|--------------------|------------------|---|
| Meaning            | JSON             | All JSON Meaning files of 8.2.4                                 |
| Recognised Text    | Unicode          | All Text files of 8.2.1   |
| Input Emotion      | JSON             | All JSON Emotion files of 8.2.3                                 |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>  |
| Machine Text       | Unicode          | Text file shall conform with [5]                                |
| Machine Emotion    | JSON             | JSON File shall validate against <a href="#">Emotion</a> Schema |

The two attributes emotion\_Name and emotion\_SetName must be present in the output JSON file of Emotion. The value of either of the two attributes may be null.

### 5.3.8 Text-To-Speech (Emotion)

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b> |
|--------------------|------------------|---------------------------------------|
| Machine Text       | Unicode          | All Text files of 8.2.1               |
| Machine Emotion    | JSON             | All JSON Emotion files of 8.2.3       |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>                    |
| Machine Speech     | .wav             | Speech file shall conform with [8]    |

### 5.3.9 Video Lip Animation

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b> |
|--------------------|------------------|---------------------------------------|
| Machine Speech     | .wav             | All Text files of 8.2.1               |
| Machine Emotion    | JSON             | All JSON Emotion files of 8.2.3       |
| Video of Face      | AVC              | All Video files of 8.2.2              |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>                    |
| Machine Face       | AVC              | Video file shall conform with [7]     |

## 6. Multimodal Question Answering (MQA)

### 6.1 Reference Model

Figure 1 reproduces the Multimodal Question Answering Reference Model.

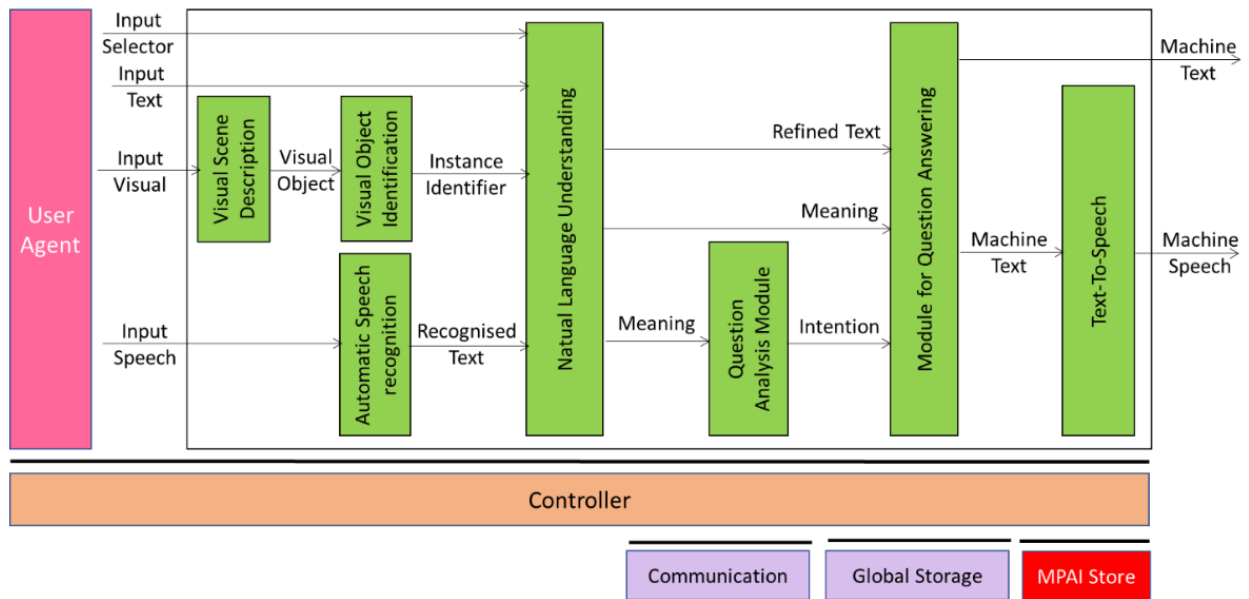


Figure 1 – Multimodal Question Answering Reference Model

**Important note.** This Conformance Testing Specification only provide methods and datasets to Test the Conformance of the Composite AIM including Visual Scene Description and Visual Object Identification Basic AIMs, not of the individual Basic AIMs..

## 6.2 AIW Conformance Testing

| Input Data     | Data Type   | Input Conformance Testing Data            |
|----------------|-------------|---|
| Input Selector | Binary data | For Input Selector=0 and Input Selector=1 |
| Text Object    | Unicode     | All Text files of 8.3.1.                  |
| Speech Object  | .wav        | All Speech files of 8.3.2.                |
| Input Image    | JPEG        | All Image files of 8.3.3.                 |
| Output Data    | Data Type   | Conformance Test                          |
| Machine Text   | Unicode     | All Text files shall conform with [5]     |
| Machine Speech | .wav        | All Speech files shall conform with [8]   |

## 6.3 AIM Conformance Testing

Table 4 gives the AIMs and the input/output data of the Multimodal Question Answering Use Case.

Table 4 - Multimodal Question Answering AIMs and I/O data

| AIM  | Input Data   | Output Data                |
|--|--|----------------------------|
| <b>Visual Scene Description + Visual Object Identification</b> | Image  | Object Instance Identifier |
| <b>Automatic Speech Recognition</b>                            | Input Speech                                       | Recognised Text            |
| <b>Natural Language Understanding</b>                          | Input Selection<br>Input Text<br>Recognised Text   | Refined Text<br>Meaning    |
| <b>Question Analysis Module</b>                                | Meaning  | Intention                  |
| <b>Answer to Question Module</b>                               | Refined Text<br>Input Text<br>Meaning<br>Intention | Output Text                |

|                       |             |               |
|-----------------------|-------------|---------------|
| <b>Text-To-Speech</b> | Output Text | Output Speech |
|-----------------------|-------------|---------------|

The input and output data of the AIMs are specified in the following paragraphs.

### 6.3.1 Visual Scene Description + Visual Object Identification

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>      |
|--------------------|------------------|--|
| Input Image        | JPEG             | All Image files of 8.3.3                   |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>                         |
| Object Instance ID | ID               | <a href="#">Instance Identifier</a> of [3] |

The InstanceLabel, LabelConfidenceLevel, Classification, ClassificationConfidenceLevel must be present in the output JSON file of InstanceIdentifier. The value of any of the four attributes may be null.

### 6.3.2 Automatic Speech Recognition

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b> |
|--------------------|------------------|---------------------------------------|
| Speech Object      | .wav             | All Speech files of 8.3.2.            |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>                    |
| Recognised Text    | Unicode          | Text files shall conform with [5]     |

### 6.3.3 Natural Language Understanding

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                                |
|--------------------|------------------|--|
| Input Selector     | Binary data      | For Input Selector=0 and Input Selector=1                            |
| Text Object        | Unicode          | All Text files of 8.3.1  |
| Recognised Text    | Unicode          | All Text files of 8.3.1  |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>   |
| Meaning            | JSON             | All JSON files shall validate against <a href="#">Meaning</a> Schema |
| Refined Text       | Unicode          | All Text files shall conform with [5]                                |

The four taggings: POS\_tagging, NE\_tagging, dependency\_tagging, and SRL\_tagging must be present in the output JSON file of Meaning. Any of the four tagging values may be null.

### 6.3.4 Question Analysis Module

| <b>Input Data</b>  | <b>Data Type</b> | <b>Input Conformance Testing Data</b>                            |
|--------------------|------------------|--|
| Meaning            | .wav             | All Meaning files of 8.3.4                                       |
| <b>Output Data</b> | <b>Data Type</b> | <b>Data Format</b>   |
| Intention          | Unicode          | All JSON files shall validate against the Intention JSON Schema. |

qtopic, qfocus, qLAT, qSAT, and qdo-main must be present in the output JSON file of Intention. The value of any of the five attributes may be null.

### 6.3.5 Answer to Question Module

| <b>Input Data</b> | <b>Data Type</b> | <b>Input Conformance Testing Data</b> |
|-------------------|------------------|---------------------------------------|
| Text              | Unicode          | All Text files of 8.3.1.              |
| Refined Text      | Unicode          | All Text files of 8.3.1.              |
| Meaning           | JSON             | All Meaning files of 8.3.4.           |
| Intention         | JSON             | All Intention files of 8.3.5.         |

| Output Data  | Data Type | Data Format                            |
|--------------|-----------|--|
| Machine Text | Unicode   | All Text files shall conform with [5]. |

### 6.3.6 Text-To-Speech

| Input Data     | Data Type | Input Conformance Testing Data           |
|----------------|-----------|--|
| Machine Text   | Unicode   | All Text files of 8.3.1.                 |
| Output Data    | Data Type | Data Format                              |
| Machine Speech | .wav      | All Speech files shall conform with [8]. |

## 7. Unidirectional Speech Translation (UST)

### 7.1 Reference Model

Figure 3 reproduces the Unidirectional Speech Translation Reference Model.

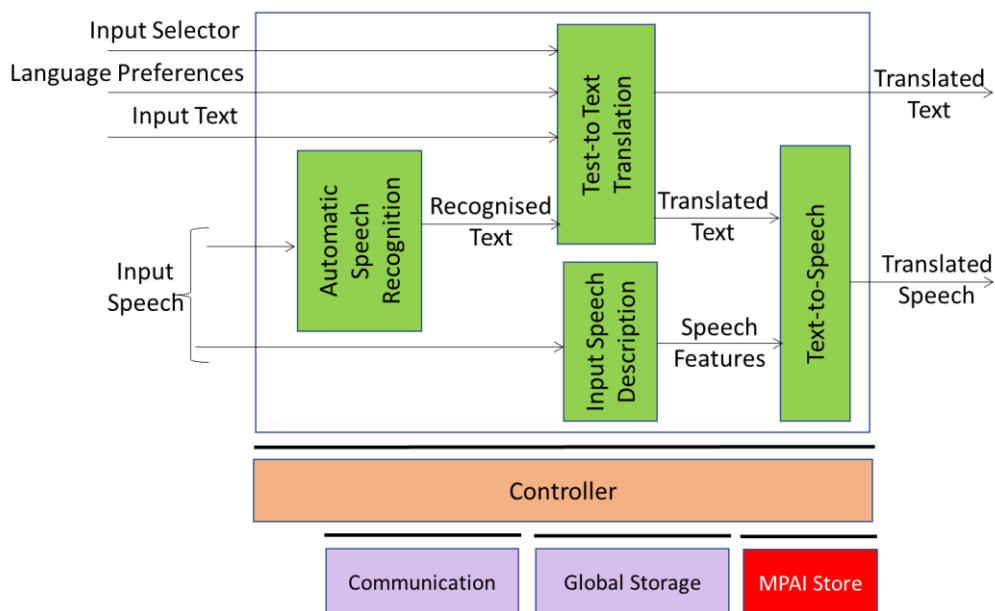


Figure 3 - Unidirectional Speech Translation Reference Model

**Important note.** This Conformance Testing Specification does not provide methods and datasets to Test the Conformance of the individual Speech Feature Extraction and Text-To-Speech Basic AIMS, only of their Composite AIMS.

### 7.2 AIW Conformance Testing

| Input Data         | Data Type   | Input Conformance Testing Data                                       |
|--------------------|-------------|--|
| Input Selection    | Binary data | For Input Selector=0 Text only.<br>For Input Selector=1 Speech only. |
| Requested Language | Unicode     | One of the languages in [4].   |
| Input Text         | .wav        | All Text files of 8.4.1.   |
| Input Speech       | JPEG        | All Speech files of 8.4.2.   |
| Output Data        | Data Type   | Conformance Test   |
| Machine Text       | Unicode     | All Text files shall conform with [5]                                |
| Machine Speech     | .wav        | All Speech files shall conform with [8]                              |

### 7.3 AIM Conformance Testing

Table 1 gives the AIMs and the input/output data of the Unidirectional Speech Translation Use Case.

Table 1 – Unidirectional Speech Translation AIMs and data formats

| AIM                                 | Input Data  | Output Data       |
|-------------------------------------|---|-------------------|
| <b>Automatic Speech Recognition</b> | Input Speech  | Recognised Text   |
| <b>Text-to-Text Translation</b>     | Input Selection<br>Requested Language<br>Input Text | Translated Text   |
| <b>Text-To-Speech</b>               | Translated Text                                     | Translated Speech |

The input and output data of the AIMs are specified in the following paragraphs.

#### 7.3.1 Speech Recognition

| Input Data     | Data Type | Input Conformance Testing Data           |
|----------------|-----------|--|
| Input Speech   | Speech    | All Speech files of 8.2.2                |
| Output Data    | Data Type | Data Format                              |
| Machine Speech | .wav      | All Speech files shall conform with [8]. |

#### 7.3.2 Text-to-Text Translation

| Input Data      | Data Type | Input Conformance Testing Data         |
|-----------------|-----------|--|
| Input Text      | Unicode   | All Text files of 8.2.1                |
| Output Data     | Data Type | Data Format                            |
| Translated Text | Unicode   | All Text files shall conform with [5]. |

#### 7.3.3 Speech Feature Extraction and Text-To-Speech

| Input Data        | Data Type | Input Conformance Testing Data           |
|-------------------|-----------|--|
| Input Speech      | Speech    | All Speech files of 8.2.2                |
| Output Data       | Data Type | Data Format                              |
| Translated Speech | Speech    | All Speech files shall conform with [8]. |

#### 7.3.4 Text-To-Speech

| Input Data        | Data Type | Input Conformance Testing Data           |
|-------------------|-----------|--|
| Translated Text   | Speech    | All Text files of 8.2.1                  |
| Output Data       | Data Type | Data Format                              |
| Translated Speech | Speech    | All Speech files shall conform with [8]. |

## 8. Data Sets

### 8.1 Introduction

Testing the Conformance of MMC-CWE, MMC-MQA, and MMC-UST requires datasets to test Data, AIMs, and AIWs. The Data Formats belong to one of Text, Audio, Video, and JSON and should have the characteristics of Table 5:

Table 5 – Data Types for Conformance Testing of MMC-CWE, MMC-MQA, and MMC-UST

| Data Type   | Characteristics  |
|-------------|--|
| Text        | The texts files are composed of Unicode characters.                  |
| Speech file | The speech files are conforming .wav files.                          |
| Video file  | The video files are conforming MP4 files [6].                        |
| JSON data   | The Emotion, Meaning, and Intention files are conforming JSON files. |

Conformance Testing may be carried out using visual and auditory inspection of a human. Appropriate software may replace human a Conformance Tester. Conformance Testing Datasets are publicly [available](#) upon registration.

## 8.2 Conversation with Emotion

### 8.2.1 Text

#### 8.2.1.1 Coherent scenarios

|         |  |
|---------|--|
| Happy   | 1. Today was a wonderful day. I spent quality time with my parents, and the restaurant was excellent as well. I look forward to seeing them again!   |
|         | 2. I'm so excited about Christmas. This year, my girlfriend and I are going to celebrate the holiday together. We'll decorate our room, and it'll be so much fun.  |
|         | 3. Today I watched a movie called 'The Pianist.' Not only was it touching, but also very absorbing. Now I feel very happy thanks to the memorable experience.  |
|         | 4. The weather is awesome these days. It is not too cold, not too hot, and the sun shines beautifully. I look forward to the picnic that is scheduled this weekend.  |
|         | 5. Nowadays my business is running very smoothly. There are no unexpected issues arising, and my employees are working very diligently. I am very relieved.  |
| Angry   | 1. Today my coworkers treated me really badly. They blamed me for the things that were neither my responsibility nor the result of my actions. This is so unfair.  |
|         | 2. I am angry with my sister. She not only does not finish her chores, but forces me to do the chores for her. This is not a new occasion, but this time I can't, stand it.  |
|         | 3. Yesterday I had an argument with a friend of mine. He always wants me to listen to him very carefully and provide advice, but when I'm in need of the help of the same sort, he doesn't fulfill his duty at all. I'm furious about this.                |
|         | 4. These days consumer price is skyrocketing. However, the government and political parties are busy blaming the external variables, not trying hard to solve the problem that ordinary citizens are facing. Why is there no one trying to be responsible? |
|         | 5. Because of my superior in my workplace, I am doing monotonous tasks all day long these days. I have to look at thousands of boring images and classify them each day, which drives me crazy. I cannot but blame my superior.                            |
| Neutral | 1. Seoul is the capital city of the Republic of Korea. It is a city of almost ten million residents. According to "The Global Livability Index" Seoul is ranked the fourth most livable city in Asia as of 2023.   |
|         | 2. There is a famous proverb, "Honesty is the best policy." In essence, it suggests that honesty is the most effective and beneficial approach in various aspects of life.   |



|  |   |
|--|---|
|  | 3. There is a famous saying, “Don’t judge a book by its cover.” This advises people not to form an opinion or make assumptions about someone or something based solely on its outward appearance.   |
|  | 4. Global warming refers to the long-term increase in Earth's average surface temperature due to human activities, primarily the emission of greenhouse gases. Greenhouse gases trap heat in the Earth's atmosphere, leading to the warming effect. |
|  | 5. Inflation is a general increase of the prices of goods and services in an economy. This is usually measured using the consumer price index (CPI).  |

### 8.2.1.2 Incoherent scenarios

| Text    | Meaning | Speech  | Face    | Sentences  |
|---------|---------|---------|---------|--|
| Happy   | Happy   | Angry   | Angry   | I’m headed to a yoga class now, and then I have a cozy evening planned with a good book. Life is good, for sure.   |
| Happy   | Happy   | Neutral | Neutral | With a big scoop of ice cream in hand, I laughed and played in the park, feeling super happy as the sun shone brightly overhead.   |
| Angry   | Angry   | Happy   | Happy   | Witnessing my neighbor being rude and disrespectful to an old stranger asking for directions, I couldn’t be sane, because that old man was my father.  |
| Neutral | Neutral | Happy   | Happy   | A political party is an organization that coordinates candidates to compete in a particular country's elections. It is common for the members of a party to hold similar ideas about politics. |
| Neutral | Neutral | Angry   | Angry   | According to Max Weber, a state is a compulsory political organization with a centralized government that maintains a monopoly of the legitimate use of force within a certain territory.      |

## 8.2.2 Audio and Video Files

### 8.2.2.1 Neutral

MPAI emotions neutral 1 audio.240309.1041.wav  
 MPAI emotions neutral 1 video.240309.1041.mp4  
 MPAI emotions neutral 1.240309.1041.mp4  
 MPAI emotions neutral 2 audio.240309.1041.wav  
 MPAI emotions neutral 2 video.240309.1041.mp4  
 MPAI emotions neutral 2.240309.1041.mp4  
 MPAI emotions neutral 3 audio.240309.1041.wav  
 MPAI emotions neutral 3 video.240309.1041.mp4  
 MPAI emotions neutral 3.240309.1041.mp4  
 MPAI emotions neutral 4 audio.240309.1041.wav  
 MPAI emotions neutral 4 video.240309.1041.mp4  
 MPAI emotions neutral 4.240309.1041.mp4  
 MPAI emotions neutral 5 audio.240309.1041.wav  
 MPAI emotions neutral 5 video.240309.1041.mp4  
 MPAI emotions neutral 5.240309.1041.mp4

### **8.2.2.2 Angry**

MPAI emotions angry 5.240309.1041.mp4  
MPAI emotions angry 5 video.240309.1041.mp4  
MPAI emotions angry 5 audio.240309.1041.wav  
MPAI emotions angry 4.240309.1041.mp4  
MPAI emotions angry 4 audio.240309.1041.wav  
MPAI emotions angry 3.240309.1041.mp4  
MPAI emotions angry 3 video.240309.1041.mp4  
MPAI emotions angry 3 audio.240309.1041.wav  
MPAI emotions angry 2.240309.1041.mp4  
MPAI emotions angry 2 video.240309.1041.mp4  
MPAI emotions angry 2 audio.240309.1041.wav  
MPAI emotions angry 1.240309.1041.mp4  
MPAI emotions angry 1 video.240309.1041.mp4  
MPAI emotions angry 1 audio.240309.1041.wav

### **8.2.2.3 Happy**

MPAI emotions happy 1 audio.240309.1041.wav  
MPAI emotions happy 1 video.240309.1041.mp4  
MPAI emotions happy 1.240309.1041.mp4  
MPAI emotions happy 2 audio.240309.1041.wav  
MPAI emotions happy 2 video.240309.1041.mp4  
MPAI emotions happy 2.240309.1041.mp4  
MPAI emotions happy 3 audio.240309.1041.wav  
MPAI emotions happy 3 video.240309.1041.mp4  
MPAI emotions happy 3.240309.1041.mp4  
MPAI emotions happy 4 audio.240309.1041.wav  
MPAI emotions happy 4 video.240309.1041.mp4  
MPAI emotions happy 4.240309.1041.mp4  
MPAI emotions happy 5 audio.240309.1041.wav  
MPAI emotions happy 5 video.240309.1041.mp4  
MPAI emotions happy 5.240309.1041.mp4

### **8.2.2.4 Incoherent**

MPAI emotions angry text happy voice.240309.1041.mp4  
MPAI emotions angry text happy voice audio.240309.1041.wav  
MPAI emotions angry text happy voice video.240309.1041.mp4  
MPAI emotions happy text angry voice.240309.1041.mp4  
MPAI emotions happy text angry voice audio.240309.1041.wav  
MPAI emotions happy text angry voice video.240309.1041.mp4  
MPAI emotions happy text neutral voice.240309.1041.mp4  
MPAI emotions happy text neutral voice audio.240309.1041.wav  
MPAI emotions happy text neutral voice video.240309.1041.mp4  
MPAI emotions neutral text angry voice.240311.0915.mp4  
MPAI emotions neutral text angry voice audio.240311.0915.wav  
MPAI emotions neutral text angry voice video.240311.0915.mp4  
MPAI emotions neutral text happy voice audio.240309.1041.wav  
MPAI emotions neutral text happy voice video.240309.1041.mp4

### 8.2.3 Emotion JSON Files

The JSON files below represent Happy, Angry, and Neutral Emotions.

```
{
  "EmotionType":{
    "emotionDegree":"high",
    "emotionName":"happy",
    "emotionSetName":"MPAI Basic Emotion Set"
  }
}

{
  "EmotionType":{
    "emotionDegree":"high",
    "emotionName":"happy",
    "emotionSetName":"MPAI Basic Emotion Set"
  }
}

{
  "EmotionType":{
    "emotionDegree":"high",
    "emotionName":"happy",
    "emotionSetName":"MPAI Basic Emotion Set"
  }
}
```

### 8.2.4 Meaning JSON Files

**Sentence 1:** Today was a wonderful day! I spent quality time with my parents, and the McDonald restaurant was excellent, too. I'm looking forward to seeing them again!

```
{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
      "POS_tagging_result": "Today/RB was/VBD a/DT wonderful/JJ day/NN !/. I/PRP spent/VBD
quality/NN time/NN with/IN my/PRP$ parents/NNS ,/, and/CC the/DT McDonald/NNP restaurant/NN
was/VBD excellent/JJ ,/, too/RB !/. I'm/NNP looking/VBG forward/RB to/TO seeing/VBG them/PRP
again/RB !/."
    },
    "NE_tagging": {
      "NE_tagging_set": "CST's named entity recogniser",
      "NE_tagging_result": " [Today,misc,uncertain] was a wonderful day ! I spent quality time
with my parents, and the [McDonald,person,likely] restaurant was excellent , too . I'm looking
forward to seeing them again!"
    },
    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": "<β>\nToday [today] <*> <atemp> ADV @ADVL #1->2\nwas [be]
<mv> V IMPF 1/3S @FS-STA #2->0\na [a] <indef> ART S @>N #3->5\nwonderful [wonderful] ADJ POS @>N
#4->5\nday [day] <dur> <per> <idf> <nhead> N S NOM @<SUBJ #5->2\n! [!] PU @PU #6->0\n</s>\n</β>\nI
[I] <*> PERS 1S NOM @SUBJ> #1->2\nspent [spend] <cjt-head> <mv> V IMPF @FS-STA #2->0\nquality
[quality] <f-q> <f-phys> <comp1> <first> <idf> <comp1> <ncomp> N S NOM @>N #3->4\ntime [time]
<ac-cat> <temp> <per> <num+> <second> <comp2> <idf> <nhead> N S NOM @<ACC #4->2\nwith [with] PRP
@<ADVL #5->2\nmy [I] <poss> <refl> <det> PERS 1S GEN @>N #6->7\nparents [parent] <Hfam> <def>
<nhead> N P NOM @P< #7->5\n, [,] PU @PU #8->0\nand [and] <clb?> <co-fin> KC @CO #9->2\nthe [the]
<def> ART S/P @>N #10->12\nMcDonald [McDonald] <*> <Proper> <first> <ncomp> N S NOM @>N
#11->12\nrestaurant [restaurant] <inst> <second> <def> <nhead> N S NOM @SUBJ> #12->13\nwas [be]
<cjt> <mv> V IMPF 1/3S @FS-STA #13->2\nexcellent [excellent] <Q:good> ADJ POS @<SC #14->13\n, [,]
PU @PU #15->0\ntoo [too] ADV @<ADVL #16->13\n. [.] PU @PU #17->0\n</s>\n</β>\nI-m [I-m] <*> <unit>
<ac-sign> <heur> <idf> <nhead> N S NOM @SUBJ> #1->2\nlooking [look] <mv> V PCP1 @ICL-ADVL
#2->0\nforward [forward] <adir> <adv1-clouse> ADV @<ADVL #3->2\nto [to] <adv1-clouse> PRP @<ADVL
#4->2\nseeing [see] <vq> <v.contact> <vtk+ADJ> <mv> V PCP1 @ICL-P< #5->4\nthem [they] PERS 3P ACC
@<ACC #6->5\nagain [again] <atemp> ADV @<ADVL #7->5\n! [!] PU @PU #8->0\n</β>"
    },
    "SRL_tagging": {
```

```

        "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
        "SRL_tagging_result": "Today/ARG1 was/PRED (a wonderful day)/ARG2 ! I/ARG0 spent/PRED
(quality time)/ARG1 (with my parents)/ARG2, and (the McDonald restaurant)/ARG1 was/PRED
excellent/ARG2, too/ARGM-ADV. I/ARG0'm looking/PRED forward/ARGM-DIR (to seeing them
again)/ARG1!"
    }
}
}

```

---

**Sentence 2:** I'm really excited about Christmas! This year, my girlfriend and I are gonna celebrate the holiday together. We're gonna decorate our room, and it'll be so much fun!

```

{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
      "POS_tagging_result": " I'm/NNP really/RB excited/VBD about/IN Christmas/NNP !/.\nThis/DT
year/NN ,/, my/PRP$ girlfriend/NN and/CC I/PRP are/VBP gon/VBG na/TO celebrate/VB the/DT
holiday/NN together/RB ./.. We're/NNP gon/VBG na/TO decorate/VB our/PRP$ room/NN ,/, and/CC
it'll/NN be/VB so/RB much/JJ fun/NN !/.."
    },
    "NE_tagging": {
      "NE_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
      "NE_tagging_result": " I'm really excited about Christmas/DATE ! This year, my girlfriend
and I are gonna celebrate the holiday together. We're gonna decorate our room, and it'll be so
much fun! "
    },
    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": "\n<\beta>\nI-m [I-m] <*> <unit> <ac-sign> <heur> <idf> <nhead> N
S NOM @NPHR #1->0\nreally [really] <ly> <ameta> <ADJ:real+ly> ADV @>A #2->3\nexcited [excited]
<np-close> ADJ POS @N< #3->1\nabout [about] <pp-temp> PRP @A< #4->3\nChristmas [Christmas] <*>
<temp> <per> <nhead> N S NOM @P< #5->4\n! [!] PU @PU #6->0\n</s>\n<\beta>\nThis [this] <*> <dem> DET
S @>N #1->2\neyear [year] <per> <dur> <def> <nhead> N S NOM @ADVL> #2->10\n, [,] PU @PU #3->0\nmy
[I] <poss> <det> PERS 1S GEN @>N #4->5\ngirlfriend [girlfriend] <cjt-head> <Hfam> <def> <nhead> N
S NOM @SUBJ> #5->8\nand [and] <co-subj> KC @CO #6->5\nI [I] <cjt> <*> PERS 1S NOM @SUBJ>
#7->5\nare [be] <vch> <aux> V PR -1/3S @FS-STA #8->0\ngonna [going=to] <complex> <aux> V PCP1
@ICL-AUX< #9->8\ncelebrate [celebrate] <mv> V INF @ICL-AUX< #10->9\nthe [the] <def> ART S/P @>N
#11->12\nholiday [holiday] <temp> <per> <def> <nhead> N S NOM @ACC #12->10\ntogether [together]
ADV @<ADVL #13->10\n. [.] PU @PU #14->0\n</s>\n<\beta>\nWe-re [We-re] <*> <Hmyth> <rem> <heur> <idf>
<nhead> N S NOM @SUBJ> #1->3\ngonna [going=to] <cjt-head> <complex> <aux> V PCP1 @FS-STA
#2->0\ndecorate [decorate] <v.contact> <mv> V INF @ICL-AUX< #3->2\nour [we] <poss> <det> PERS GEN
1P @>N #4->5\nroom [room] <Lh> <am> <def> <nhead> N S NOM @ACC #5->3\n, [,] PU @PU #6->0\nand
[and] <c1b?> KC @CO #7->5\nit-ll [it-ll] <heur> <idf> <nhead> N S NOM @SUBJ> #8->9\nbe [be] <cjt>
<mv> V SUBJ @FS-STA #9->2\nso [so] <aquant> ADV @>A #10->11\nmuch [much] <quant> DET ABS S @>N
#11->12\nfun [fun] <sem-c> <percep-f> <idf> <nhead> N S NOM @SC #12->9\n! [!] PU @PU
#13->0\n</\beta>"
    },
    "SRL_tagging": {
      "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
      "SRL_tagging_result": " I/ARG1 'm/PRED (really excited about Christmas)/ARG2! This year,
(my girlfriend and I) /ARG0 are gonna celebrate/PRED (the holiday)/ARG1 together/ARGM-MNR.
We/ARG0 're gonna decorate/PRED (our room)/ARG1, and it/ARG1 'll/ARGM-MOD be/PRED (so much
fun)/ARG2 !"
    }
  }
}

```

---

**Sentence 3:** Today I watched a movie called 'The Pianist.' It was not only touching, but really absorbing, too. Now I'm feeling really happy, thanks to this memorable experience.

```

{
  "meaning": {

```

```

"POS_tagging": {

  "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",

  "POS_tagging_result": " Today/RB I/PRP watched/VBD a/DT movie/NN called/VBN '/' The/DT
Pianist/NNP ./.' /POS It/PRP was/VBD not/RB only/RB touching/VBG ,/, but/CC really/RB
absorbing/VBG ,/, too/RB ./.\nNow/RB I'm/NNP feeling/NN really/RB happy/JJ ,/, thanks/NNS to/TO
this/DT memorable/JJ experience/NN ./.'"

},

"NE_tagging": {

  "NE_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/sr1.html",

  "NE_tagging_result": " Today I watched a movie called 'The Pianist.'/WORK_OF_ART It was not
only touching, but really absorbing, too. Now I'm feeling really happy, thanks to this memorable
experience."

},

"dependency_tagging": {

  "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",

  " dependency_tagging_result": "\n<\nToday [today] <*> <atemp> ADV @ADVL> #1->3\nI [I] <*>
PERS 1S NOM @SUBJ> #2->3\nwatched [watch] <DL:bio> <mv> V IMPF @FS-STA #3->0\na [a] <indef> ART S
@>N #4->5\nmovie [movie] <sem-w> <DL:bio> <idf> <nhead> N S NOM @<ACC #5->3\ncalled [call]
<vtk+N> <vtk+ADJ> <vtk+N> <vtk+PROP> <vq> <v.contact> <DL:bio> <mv> <np-close> V PCP2 PAS @ICL-N<
#6->5\n-The [-The] <heur> <DL:bio> <idf> <nhead> N S NOM @<SC #7->6\nPianist [Pianist] <*>
<Proper> <DL:bio> <nhead> N S NOM @<OC #8->6\n. [.] PU @PU #9->0\n<\n- [-] PU @PU
#1->0\n</\n</s>\n<\nIt [it] <*> PERS NEU 3S NOM @SUBJ> #1->2\nwas [be] <DL:bio> <mv> V IMPF
1/3S @FS-STA #2->0\nnot [not] ADV @>A #3->4\nonly [only] <ly> <ADJ:on+ly> <adv1-close> ADV @<ADVL
#4->2\ntouching [touching] <DL:bio> ADJ POS @<SC #5->2\n, [,] PU @PU #6->0\nbut [but] KC @CO
#7->5\nreally [really] <ly> <ameta> <ADJ:real+ly> ADV @ADVL> #8->9\nabsorbing [absorb]
<v.contact> <DL:bio> <mv> V PCP1 @ICL-N<PRED #9->1\n, [,] PU @PU #10->0\ntoo [too] <adv1-close>
ADV @<ADVL #11->9\n. [.] PU @PU #12->0\n</s>\n<\nNow [now] <*> <atemp> ADV @ADVL #1->0\nI-m [I-
m] <*> <unit> <ac-sign> <DL:bio> <heur> <nhead> N S NOM @NPHR #2->1\nfeeling [feel] <v.contact>
<v-cog> <DL:bio> <mv> <np-close> V PCP1 @ICL-N<PRED #3->2\nreally [really] <ly> <ameta>
<ADJ:real+ly> ADV @>A #4->5\nhappy [happy] <jpsych> <DL:bio> ADJ POS @<SC #5->3\n, [,] PU @PU
#6->0\nthanks to [thanks=to] <insertion> <complex> PRP @<ADVL #7->3\nthis [this] <dem> DET S @>N
#8->10\nmemorable [memorable] <DL:bio> ADJ POS @>N #9->10\nexperience [experience] <f-psych>
<percep-f> <DL:bio> <def> <nhead> N S NOM @P< #10->7\n. [.] PU @PU #11->0\n</\n>"

},

"SRL_tagging": {

  "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/sr1.html",

  "SRL_tagging_result": "Today/ARG-TMP I/ARG0 watched/PRED a movie/ARG1 called/PRED 'The
Pianist.' It/ARG1 was/PRED (not only touching, but really absorbing, too)/ARG2. Now/ARG-TMP
I/ARG0 'm feeling/PRED (really happy)/ARG1, thanks to this memorable experience."

}

}

}

```

## 8.3 Multimodal Question Answering

### 8.3.1 Text Files

Q1: What is the tool in the picture?

Q2: What is the nickname of the person in the picture?

- Q3: What is the job of the person on the left hand-side in the picture  
 Q4: What is the family name of the person in the centre of the picture?  
 Q5: What is the name of the square in the picture?

### 8.3.2 Audio Files

- Q1.wav  
 Q2.wav  
 Q3.wav  
 Q4.wav  
 Q5.wav

### 8.3.3 Images

|                |                                  |
|----------------|----------------------------------|
| Images for Q1  | Q1-1.jpg<br>Q1-2.jpg<br>Q1-3.jpg |
| Image for Q2   | Q2-Joseph Gordon Levitt.jpg      |
| Image for Q3   | Q3-1.jpg<br>Q3-2.jpg             |
| images for Q4  | Q4-1.jpg<br>Q4-1.jpg<br>Q4-3.jpg |
| 1 image for Q5 | Q5-1.jpg                         |

### 8.3.4 Meaning JSON Files

**Sentence 1:** What is the tool in the picture?

```
{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations),
https://cst.dk/online/pos_tagger/uk/",
      "POS_tagging_result": "What/WP is/VBZ the/DT tool/NN in/IN the/DT picture/NN ?/."
    },
    "NE_tagging": {
      "NE_tagging_set": "CST's named entity recogniser,
https://cst.dk/online/navnegenkenderCSTNER/uk/",
      "NE_tagging_result": " [What,misc,uncertain] is the tool in the picture ?"
    },
    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.vis1.dk/vis1/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": "<β>\nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2\nis
[be] <mv> V PR 3S @FS-QUE #2->0\nthe [the] <def> ART S/P @>N #3->4\ntool [tool] <tool> <def>
<nhead> N S NOM @<SUBJ #4->2\nin [in] <advl-fs> PRP @<ADVL #5->2\nthe [the] <def> ART S/P @>N
#6->7\npicture [picture] <pict> <repr> <def> <nhead> N S NOM @P< #7->5\n? [?] PU @PU #8->0\n</β>"
    },
    "SRL_tagging": {
      "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
      "SRL_tagging_result": " What/ARG2 is/PRED (the tool in the picture)/ARG1 ?"
    }
  }
}
```

**Sentence 2:** What is the nickname of the person in the picture?

What/WP is/VBZ the/DT nickname/NN of/IN the/DT person/NN in/IN the/DT picture/NN ?/.

```
{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
```

```

    "POS_tagging_result": " What/WP is/VBZ the/DT nickname/NN of/IN the/DT person/NN in/IN
the/DT picture/NN ?/."
  },
  "NE_tagging": {
    "NE_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/ner.html ",
    "NE_tagging_result": ""
  },
  "dependency_tagging": {
    "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
    "dependency_tagging_result": " <β>\nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2\nis
[be] <mv> V PR 3S @FS-QUE #2->0\nthe [the] <def> ART S/P @>N #3->4\nnickname [nickname] <ac-cat>
<def> <nhead> N S NOM @<SUBJ #4->2\nof [of] <np-close> PRP @N< #5->4\nthe [the] <def> ART S/P @>N
#6->7\nperson [person] <H> <def> <nhead> N S NOM @P< #7->5\nin [in] <advl-fs> PRP @<ADVL
#8->2\nthe [the] <def> ART S/P @>N #9->10\npicture [picture] <pict> <repr> <def> <nhead> N S NOM
@P< #10->8\n? [?] PU @PU #11->0\n</β>"
  },
  "SRL_tagging": {
    "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
    "SRL_tagging_result": " What/ARG2 is/PRED (the nickname of the person in the
picture)/ARG1 ?"
  }
}
}
}

```

**Sentence 3:** What is the job of the person on the left hand-side in the picture?

```

{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
      "POS_tagging_result": " What/WP is/VBZ the/DT job/NN of/IN the/DT person/NN on/IN the/DT
left/VBN hand-side/JJ in/IN the/DT picture/NN ?/."
    },
    "NE_tagging": {
      "NE_tagging_set": " https://cst.dk/online/navnegenkenderCSTNER/uk/",
      "NE_tagging_result": " [What,misc,uncertain] is the job of the person on the left hand-side
in the picture ?"
    },
    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": " <β>\nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2\nis
[be] <mv> V PR 3S @FS-QUE #2->0\nthe [the] <def> ART S/P @>N #3->4\njob [job] <pos-soc> <sem-c>
<def> <nhead> N S NOM @<SUBJ #4->2\nof [of] <np-close> PRP @N< #5->4\nthe [the] <def> ART S/P @>N
#6->7\nperson [person] <H> <def> <nhead> N S NOM @P< #7->5\non [on] <advl-fs> PRP @<ADVL
#8->2\nthe [the] <def> ART S/P @>N #9->11\nleft [left] ADJ POS @>N #10->11\nhand-side [hand-side]
<Lsurf> <HH> <geom> <heur> <def> <nhead> N S NOM @P< [hand-side] <heur> <def> N S NOM @P<
#11->8\nin [in] <advl-fs> PRP @<ADVL #12->2\nthe [the] <def> ART S/P @>N #13->14\npicture
[picture] <pict> <repr> <def> <nhead> N S NOM @P< #14->12\n? [?] PU @PU #15->0\n</β>"
    },
    "SRL_tagging": {
      "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
      "SRL_tagging_result": " What/ARG2 is/PRED (the job of the person on the left hand-side in
the picture)/ARG1 ?"
    }
  }
}
}

```

**Sentence 4:** What is the family name of the person in the centre of the picture?

```

{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
      "POS_tagging_result": " What/WP is/VBZ the/DT family/NN name/NN of/IN the/DT person/NN
in/IN the/DT centre/NN of/IN the/DT picture/NN ?/."
    },
    "NE_tagging": {
      "NE_tagging_set": " https://cst.dk/online/navnegenkenderCSTNER/uk/",
      "NE_tagging_result": " [What,misc,uncertain] is the family name of the person in the centre
of the picture ?"
    },
  }
}

```

```

    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": " <β>\nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2\nis
[be] <mv> V PR 3S @FS-QUE #2->0\nthe [the] <def> ART S/P @>N #3->5\nfamily [family] <HH> <comp1>
<comp1> <ncomp> N S NOM @>N #4->5\nname [name] <ac-cat> <comp2> <def> <nhead> N S NOM @<SUBJ
#5->2\nof [of] <np-close> PRP @N< #6->5\nthe [the] <def> ART S/P @>N #7->8\nperson [person] <H>
<def> <nhead> N S NOM @P< #8->6\nin [in] <adv1-fs> PRP @<ADVL #9->2\nthe [the] <def> ART S/P @>N
#10->11\ncentre [centre] <Labs> <inst> <def> <nhead> N S NOM @P< #11->9\nof [of] <np-close> PRP
@N< #12->11\nthe [the] <def> ART S/P @>N #13->14\npicture [picture] <pict> <repr> <def> <nhead> N
S NOM @P< #14->12\n? [?] PU @PU #15->0\n</β>"
    },
    "SRL_tagging": {
      "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/sr1.html",
      "SRL_tagging_result": " What/ARG2 is/PRED (the family name of the person in the centre of
the picture)/ARG1 ?"
    }
  }
}

```

**Sentence 5:** What is the name of the square in the picture?

```

{
  "meaning": {
    "POS_tagging": {
      "POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations)",
      "POS_tagging_result": " What/WP is/VBZ the/DT name/NN of/IN the/DT square/NN of/IN the/DT
picture/NN ?/."
    },
    "NE_tagging": {
      "NE_tagging_set": " https://cst.dk/online/navnegenkenderCSTNER/uk/",
      "NE_tagging_result": "[What,misc,uncertain] is the name of the square of the picture ?"
    },
    "dependency_tagging": {
      "dependency_tagging_set": "CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php ",
      "dependency_tagging_result": " \n<β>\nWhat [what] <clb> <*> <interr> INDP S/P @SC>
#1->2\nis [be] <mv> V PR 3S @FS-QUE #2->0\nthe [the] <def> ART S/P @>N #3->4\nname [name] <ac-
cat> <sem-c> <def> <nhead> N S NOM @<SUBJ #4->2\nof [of] <np-close> PRP @N< #5->4\nthe [the]
<def> ART S/P @>N #6->7\nsquare [square] <Lh> <geom> <def> <nhead> N S NOM @P< #7->5\nof [in]
<np-close> PRP @N< #8->7\nthe [the] <def> ART S/P @>N #9->10\npicture [picture] <pict> <repr>
<def> <nhead> N S NOM @P< #10->8\n? [?] PU @PU #11->0\n</β>"
    },
    "SRL_tagging": {
      "SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/sr1.html",
      "SRL_tagging_result": " What/ARG2 is/PRED (the name of the square of the picture)/ARG1 ?"
    }
  }
}

```

### 8.3.5 Intention JSON Files

Q1: What is the tool in the picture?

```

{
  "Intention": {
    "qtopic": "tool",
    "qfocus": "What",
    "qLAT": "tool",
    "qSAT": "ETC",
    "qdomain": "everyday life"
  }
}

```

Q2: What is the nickname of the person in the picture?

```

{
  "Intention": {
    "qtopic": "person",
    "qfocus": "What",
    "qLAT": "nickname",
    "qSAT": "PS_NAME",
    "qdomain": "famous people"
  }
}

```



```
}  
}
```

Q3: What is the job of the person on the left hand-side in the picture

```
{  
  "Intention": {  
    "qtopic": "person",  
    "qfocus": "What",  
    "qLAT": "job",  
    "qSAT": "CV_OCCUPATION",  
    "qdomain": "famous people"  
  }  
}
```

Q4: What is the family name of the person in the centre of the picture?

```
{  
  "Intention": {  
    "qtopic": "person",  
    "qfocus": "What",  
    "qLAT": "family name",  
    "qSAT": "PS_NAME",  
    "qdomain": "famous people"  
  }  
}
```

Q5: What is the name of the square in the picture?

```
{  
  "Intention": {  
    "qtopic": "square",  
    "qfocus": "What",  
    "qLAT": "square",  
    "qSAT": "LC_TOUR",  
    "qdomain": "traveling"  
  }  
}
```

## 8.4 Unidirectional Speech Translation

### 8.4.1 Text Files

The Text Files of Conversation with Emotion are used.

### 8.4.2 Audio Files

The Audio Files of Conversation with Emotion are used.

## Annex 1 - MPAI-wide terms and definitions

The capitalised Terms used in this standard that are not already included in *Table 1* are defined in Table 6.

NOTE: A hyphenated entry for e.g., “- Testing” should be read as adding that word to the closest non-hyphenated entry above it – in this case, “Conformance,” giving “Conformance Testing” as the complete entry name.

*Table 6 - MPAI-wide Terms*

| <b>Term</b>           | <b>Definition</b>  |
|-----------------------|--|
| Access                | Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.  |
| AI Framework (AIF)    | The environment where AIWs are executed.   |
| AI Model (AIM)        | A data processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs. |
| AI Workflow (AIW)     | A structured aggregation of AIMs implementing a Use Case receiving AIW-specific inputs and producing AIW-specific outputs according to the AIW Function.                 |
| Application Standard  | An MPAI Standard designed to enable a particular application domain.   |
| Assessment Laboratory | A laboratory accredited to Assess the Grade of Performance of Implementations.   |
| Channel               | A connection between an output port of an AIM and an input port of an AIM. The term “connection” is also used as synonymous.   |
| Communication         | The infrastructure that implements message passing between AIMs.   |
| Component             | One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent  |
| Composite AIM         | An AIM aggregating more than one AIM.  |
| Component             | One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent  |
| Conformance           | The attribute of an Implementation of being a correct technical Implementation of a Technical Specification.   |
| - Testing             | The normative document specifying the Means to Test the Conformance of an Implementation.  |
| - Testing Dataset     | A dataset used to Test the Conformance of an implementation to a Technical Specification.  |
| - Testing Means       | Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation.   |
| - Testing Procedure   | The sequence of steps to be performed to Test the Conformance of an implementation.  |
| - Testing Tools       | Devices and/or software used to Test the Conformance of an implementation.   |
| Connection            | A channel connecting an output port of an AIM and an input port of an AIM.   |

|  |  |
|--|--|
| Controller                                   | A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed  |
| Data   | Information in digital form.   |
| - Format                                     | The standard digital representation of Data.   |
| - Type                                       | An instance of Data with a specific Data Format.   |
| - Semantics                                  | The meaning of Data.   |
| Descriptor                                   | Coded representation of a text, audio, speech, or visual feature.  |
| Digital Representation                       | Data corresponding to and representing a physical entity.  |
| Ecosystem                                    | The ensemble of actors making it possible for a User to execute an application composed of an AIF, one or more AIWs, each with one or more AIMs potentially sourced from independent implementers.   |
| Explainability                               | The ability to trace the output of an Implementation back to the inputs that have produced it.   |
| Fairness                                     | The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results.   |
| Function                                     | The operations effected by an AIW or an AIM on input data.   |
| Global Storage                               | A Component to store data shared by AIMs.  |
| AIM/AIW Storage                              | A Component to store data of the individual AIMs.  |
| Identifier                                   | A name that uniquely identifies an Implementation.   |
| Implementation                               | <ol style="list-style-type: none"> <li>1. An embodiment of the MPAI-AIF Technical Specification, or</li> <li>2. An AIW or AIM of a particular Level (1-2-3) conforming with a Use Case of an MPAI Application Standard.</li> </ol>   |
| Implementer                                  | A legal entity implementing MPAI Technical Specifications.   |
| ImplementerID (IID)                          | A unique name assigned by the ImplementerID Registration Authority to an Implementer.  |
| ImplementerID Registration Authority (IIDRA) | The entity appointed by MPAI to assign ImplementerID's to Implementers.  |
| Instance ID                                  | Instance of a class of Objects and the Group of Objects the Instance belongs to.   |
| Interoperability                             | The ability to functionally replace an AIM with another AIW having the same Interoperability Level   |
| - Level                                      | <p>The attribute of an AIW and its AIMs to be executable in an AIF Implementation and to:</p> <ol style="list-style-type: none"> <li>1. Be proprietary (Level 1)</li> <li>2. Pass the Conformance Testing (Level 2) of an Application Standard</li> <li>3. Pass the Performance Testing (Level 3) of an Application Standard.</li> </ol> |
| Knowledge Base                               | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces   |
| Message                                      | A sequence of Records transported by Communication through Channels.   |
| Normativity                                  | The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard.  |
| Performance                                  | The attribute of an Implementation of being Reliable, Robust, Fair and Replicable.   |

|                         |   |
|-------------------------|---|
| - Assessment            | The normative document specifying the Means to Assess the Grade of Performance of an Implementation.  |
| - Assessment Means      | Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation.  |
| - Assessor              | An entity Assessing the Performance of an Implementation.   |
| Profile                 | A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset.  |
| Record                  | A data structure with a specified structure   |
| Reference Model         | The AIMs and their Connections in an AIW.   |
| Reference Software      | A technically correct software implementation of a Technical Specification containing source code, or source and compiled code.   |
| Reliability             | The attribute of an Implementation that performs as specified by the Application Standard, profile, and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer.  |
| Replicability           | The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor.  |
| Robustness              | The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence.   |
| Scope                   | The domain of applicability of an MPAI Application Standard   |
| Service Provider        | An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users.  |
| Standard                | A set of Technical Specification, Reference Software, Conformance Testing, Performance Assessment, and Technical Report of an MPAI application Standard.  |
| Technical Specification | (Framework) the normative specification of the AIF.<br>(Application) the normative specification of the set of AIWs belonging to an application domain along with the AIMs required to Implement the AIWs that includes:<br>1. The formats of the Input/Output data of the AIWs implementing the AIWs.<br>2. The Connections of the AIMs of the AIW.<br>3. The formats of the Input/Output data of the AIMs belonging to the AIW. |
| Time Base               | The protocol specifying how Components can access timing information  |
| Topology                | The set of AIM Connections of an AIW.   |
| Use Case                | A particular instance of the Application domain target of an Application Standard.  |
| User                    | A user of an Implementation.  |
| User Agent              | The Component interfacing the user with an AIF through the Controller   |
| Version                 | A revision or extension of a Standard or of one of its elements.  |
| Zero Trust              | A cybersecurity model primarily focused on data and service protection that assumes no implicit trust.  |

## **Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative)**

The notices and legal disclaimers given below shall be borne in mind when [downloading](#) and using approved MPAI Standards.

In the following, “Standard” means the collection of four MPAI-approved and [published](#) documents: “Technical Specification”, “Reference Software” and “Conformance Testing” and, where applicable, “Performance Testing”.

### Life cycle of MPAI Standards

MPAI Standards are developed in accordance with the [MPAI Statutes](#). An MPAI Standard may only be developed when a Framework Licence has been adopted. MPAI Standards are developed by especially established MPAI Development Committees who operate on the basis of consensus, as specified in Annex 1 of the [MPAI Statutes](#). While the MPAI General Assembly and the Board of Directors administer the process of the said Annex 1, MPAI does not independently evaluate, test, or verify the accuracy of any of the information or the suitability of any of the technology choices made in its Standards.

MPAI Standards may be modified at any time by corrigenda or new editions. A new edition, however, may not necessarily replace an existing MPAI standard. Visit the [web page](#) to determine the status of any given published MPAI Standard.

Comments on MPAI Standards are welcome from any interested parties, whether MPAI members or not. Comments shall mandatorily include the name and the version of the MPAI Standard and, if applicable, the specific page or line the comment applies to. Comments should be sent to the [MPAI Secretariat](#). Comments will be reviewed by the appropriate committee for their technical relevance. However, MPAI does not provide interpretation, consulting information, or advice on MPAI Standards. Interested parties are invited to join MPAI so that they can attend the relevant Development Committees.

### Coverage and Applicability of MPAI Standards

MPAI makes no warranties or representations of any kind concerning its Standards, and expressly disclaims all warranties, expressed or implied, concerning any of its Standards, including but not limited to the warranties of merchantability, fitness for a particular purpose, non-infringement etc. MPAI Standards are supplied “AS IS”.

The existence of an MPAI Standard does not imply that there are no other ways to produce and distribute products and services in the scope of the Standard. Technical progress may render the technologies included in the MPAI Standard obsolete by the time the Standard is used, especially in a field as dynamic as AI. Therefore, those looking for standards in the Data Compression by Artificial Intelligence area should carefully assess the suitability of MPAI Standards for their needs.

**IN NO EVENT SHALL MPAI BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND**

ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

MPAI alerts users that practicing its Standards may infringe patents and other rights of third parties. Submitters of technologies to this standard have agreed to licence their Intellectual Property according to their respective Framework Licences.

Users of MPAI Standards should consider all applicable laws and regulations when using an MPAI Standard. The validity of Conformance Testing is strictly technical and refers to the correct implementation of the MPAI Standard. Moreover, positive Performance Assessment of an implementation applies exclusively in the context of the [MPAI Governance](#) and does not imply compliance with any regulatory requirements in the context of any jurisdiction. Therefore, it is the responsibility of the MPAI Standard implementer to observe or refer to the applicable regulatory requirements. By publishing an MPAI Standard, MPAI does not intend to promote actions that are not in compliance with applicable laws, and the Standard shall not be construed as doing so. In particular, users should evaluate MPAI Standards from the viewpoint of data privacy and data ownership in the context of their jurisdictions.

Implementers and users of MPAI Standards documents are responsible for determining and complying with all appropriate safety, security, environmental and health and all applicable laws and regulations.

#### Copyright

MPAI draft and approved standards, whether they are in the form of documents or as web pages or otherwise, are copyrighted by MPAI under Swiss and international copyright laws. MPAI Standards are made available and may be used for a wide variety of public and private uses, e.g., implementation, use and reference, in laws and regulations and standardisation. By making these documents available for these and other uses, however, MPAI does not waive any rights in copyright to its Standards. For inquiries regarding the copyright of MPAI standards, please contact the [MPAI Secretariat](#).

The Reference Software of an MPAI Standard is released with the [MPAI Modified Berkeley Software Distribution licence](#). However, implementers should be aware that the Reference Software of an MPAI Standard may reference some third party software that may have a different licence.

## Annex 3 - The Governance of the MPAI Ecosystem (Informative)

### Level 1 Interoperability

With reference to **Error! Reference source not found.**, MPAI issues and maintains a standard – called MPAI-AIF – whose components are:

1. An environment called AI Framework (AIF) running AI Workflows (AIW) composed of inter-connected AI Modules (AIM) exposing standard interfaces.
2. A distribution system of AIW and AIM Implementation called MPAI Store from which an AIF Implementation can download AIWs and AIMs.

A Level 1 Implementation shall implement the MPAI-AIF Technical Specification executing AIWs composed of AIMs able to call the MPAI-AIF APIs.

|                        |   |
|------------------------|---|
| Implementers' benefits | Upload to the MPAI Store and have globally distributed Implementations of |
|                        | - AIFs conforming to MPAI-AIF.  |
|                        | - AIWs and AIMs performing proprietary functions executable in AIF.       |
| Users' benefits        | Rely on Implementations that have been tested for security.               |
| MPAI Store's role      | - Tests the Conformance of Implementations to MPAI-AIF.                   |
|                        | - Verifies Implementations' security, e.g., absence of malware.           |
|                        | - Indicates unambiguously that Implementations are Level 1.               |

### Level 2 Interoperability

In a Level 2 Implementation, the AIW must be an Implementation of an MPAI Use Case and the AIMs must conform with an MPAI Application Standard.

|                        |  |
|------------------------|--|
| Implementers' benefits | Upload to the MPAI Store and have globally distributed Implementations of  |
|                        | - AIFs conforming to MPAI-AIF.   |
|                        | - AIWs and AIMs conforming to MPAI Application Standards.  |
| Users' benefits        | - Rely on Implementations of AIWs and AIMs whose Functions have been reviewed during standardisation.            |
|                        | - Have a degree of Explainability of the AIW operation because the AIM Functions and the data Formats are known. |
| Market's benefits      | - Open AIW and AIM markets foster competition leading to better products.  |
|                        | - Competition of AIW and AIM Implementations fosters AI innovation.  |
| MPAI Store's role      | - Tests Conformance of Implementations with the relevant MPAI Standard.  |
|                        | - Verifies Implementations' security.  |
|                        | - Indicates unambiguously that Implementations are Level 2.  |

### Level 3 Interoperability

MPAI does not generally set standards on how and with what data an AIM should be trained. This is an important differentiator that promotes competition leading to better solutions. However, the performance of an AIM is typically higher if the data used for training are in greater quantity and more in tune with the scope. Training data that have large variety and cover the spectrum of all cases of interest in breadth and depth typically lead to Implementations of higher “quality”.

For Level 3, MPAI normatively specifies the process, the tools and the data or the characteristics of the data to be used to Assess the Grade of Performance of an AIM or an AIW.

Implementers' benefits May claim their Implementations have passed Performance Assessment.

Users' benefits Get assurance that the Implementation being used performs correctly, e.g., it has been properly trained.

Market's benefits Implementations' Performance Grades stimulate the development of more Performing AIM and AIW Implementations.

MPAI Store's role - Verifies the Implementations' security  
- Indicates unambiguously that Implementations are Level 3.

### The MPAI ecosystem

Figure 4 is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations as specified in the Governance of the MPAI Ecosystem Specification [11]:

1. MPAI establishes and controls the not-for-profit MPAI Store.
2. MPAI appoints Performance Assessors.
3. MPAI publishes Standards.
4. Implementers submit Implementations to Performance Assessors.
5. If the Implementation Performance is acceptable, Performance Assessors inform Implementers and the MPAI Store.
6. Implementers submit Implementations to the MPAI Store tested for Conformance and security.
7. Users download and use Implementations and submit experience scores.

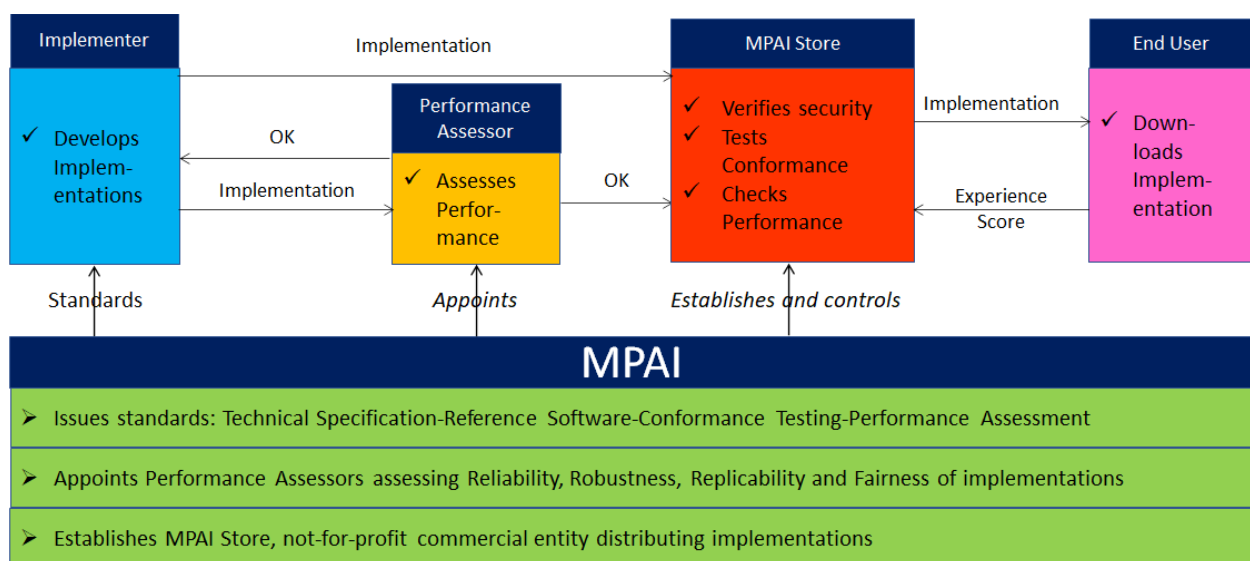


Figure 4 - The MPAI ecosystem operation